DOCUMENT RESUME

ED 293 865 TM 011 440

AUTHOR Rudner, Lawrence M.; Eissenberg, Thomas E.
TITLE Standard Setting Practices for Teacher Tests.
PUB DATE Apr 88
NOTE 13p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (New Orleans, LA, April 6-8, 1988).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Cutting Scores; Minimum Competency Testing; *Pass Fail Grading; *Scoring; *State Standards; *Teacher Certification; Teacher Qualifications; Testing Programs
IDENTIFIERS *NTE Core Battery; *Standard Setting; Teacher Competency Testing

ABSTRACT

Standard setting practices in states using the NTE (formerly the National Teacher Examinations) were examined for 1987. The NTE is composed of two segments: (1) a Core Battery covering the communication skills and general knowledge skills; and (2) a test of professional knowledge about teaching. The processes used to establish passing scores for teacher tests, recommended study scores, scores established by the states, and implications of the passing scores are discussed. Each state determined the level of performance expected of a minimally qualified applicant for certification through establishment of a "study score" that a minimally qualified individual would obtain if a test were perfectly valid. The computed passing score takes errors of measurement into account. Twelve of 14 states using the NTE Core Battery adjusted passing scores below study scores. Passing scores ranged from 630 to 657, averaging 8 points below study scores. The average passing score on the professional knowledge examination of the NTE is 47 of 104 items, which is scarcely higher than chance. Practical considerations of candidate availability seem to preclude higher standards, but there is an apparent conflict between the stated goals of teacher certification testing programs and current passing standards. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED293865

Standard Setting Practices

for

Teacher Tests

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
LAWRENCE M. RUDNER
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Lawrence M. Rudner
LMP Associates, Chevy Chase, Maryland
American Institutes for Research, Washington, DC

Thomas E. Eissenberg
American Institutes for Research, Washington, DC

April 1988

Paper presented at the annual meeting of the National Council on Measurement In Education, New Orleans, LA.

TM 011 440

Standard Setting Practices for Teacher Tests

Lobbying for a teacher testing program, one governor referred to "the contribution the testing process would make in our efforts to restore the teaching profession to the position of public trust and esteem it deserves." The board of education president in another state proclaimed that, "we are trying to assure the public that we have better quality applicants coming into the teaching profession." One legislator in yet another state remarked, "Passing scores should be based on what is needed to perform the job, regardless of how many pass or fail."

These officials and other advocates of state teacher testing programs see test-enforced standards as means to screen out unqualified individuals, to strengthen the teaching profession, and to attract better qualified candidates. As a result of these programs, the public's confidence in teachers, teaching, and the schools is expected to improve.

In order to meet these goals, there must be a sufficient supply of prospective teachers, the test must measure appropriate content, and the test standards must be sufficiently high. Whether current teacher testing programs establish meaningful standards, however, has been an issue of debate. Some accuse teacher certification examinations of ensuring that new teachers meet "only the most minimum standards of academic ability." Such accusations suggest that teacher tests may not be rigorous enough to be effective.

This paper examines the standard setting practices in states using the NTE examination. The processes used to establish passing scores for teacher tests, recommended study scores, scores established by the states, and the implications of the passing scores are discussed.

Data Sources

A variety of data sources were used in this study. Study score data was extracted from an excellent validation study conducted for the state of Montana (Zetler, 1986) and from a survey conducted by the Office of Educational Research and Improvement (Rudner, 1987). Established passing scores and passing rate data was extracted from these documents and a report from the American Association for Colleges of Teacher Education (1986). Data needed to compute the distribution of NTE test scores came from an information flier published by the NTE programs (NTE, 1984). Data from these reports were verified and updated in preparation of the annual ERIC/TM digest on teacher testing (Eissenberg and Rudner, 1988).

Examined tests

While a wide range of instruments are used in teacher testing, this study concentrated on teacher certification testing using the Core Battery of the NTE examination. These NTE tests are the most frequently used teacher certification instruments. A total of 34 states were scheduled to begin implementation of teacher certification testing by the end of 1987. Of these states with teacher certification tests, 13 had been using the NTE and another two were beginning to use the NTE in 1987 (Rudner, 1988).

Formerly called the National Teacher Examinations, the NTE is composed of a Core Battery covering the communication skills of listening reading and writing; the general knowledge skills of social studies, mathematics, literature, fine arts, and science; and a test of professional knowledge about teaching. The complete battery contains 340 multiple choice questions and one essay item. It requires 5.5 hours to complete. Subject matter tests in 26 fields are also available.

Standard Setting Procedures

Each state must determine the level of performance it expects of a minimally qualified applicant for certification. The process involves gathering and analyzing judgments made by experienced teachers, administrators, and teacher educators within the state. These judgments are then combined to form

a "study score" or the score that the judges feel a minimally qualified individual would obtain if the test were perfectly valid. Since the tests are not perfectly valid, the State Department of Education then takes the errors of measurement into account and establishes a passing score which is different, almost always lower, than the study score.

There are several approaches to quantifying judgments as part of the standard-setting process. Typically, the first step in the process is developing a hypothetical reference group of minimally qualified individuals just graduating from teacher preparation programs.

With this hypothetical reference group in mind, panel members then estimate the percent of individuals in the group that would be able to correctly answer each question. The average estimated percents of minimally qualified people that would answer correctly are then added to determine the study score for the test.

The standard setting process is not without problems and limitations. It attempts to make a judgmental process systematic. But the process remains judgmental, and, as such has several problems and limitations.

Because of imprecision in estimating the study score and in measuring a candidate's ability, minimally qualified candidates do not necessarily obtain scores above the study score. If the study score, which represents the score that would be obtained if the test were perfectly valid, were adopted as the passing score, then some qualified candidates would most likely fail the test and be improperly denied certification.

State Departments of Education are left with making a difficult decision: how should errors in measurement be treated in determining the passing score? Because of errors in measurement, there will always be 4 groups of examinees:

1) true positives -- individuals who have the ability and have a test score above the passing score

2) true negatives -- individuals who do not have the ability and have a test score below the passing score
3) false positives -- individuals who do not have the ability, but have a test score above the passing score
4) false negative -- individuals who have the ability but have a test score below the passing score.

The state must evaluate the consequences of group 3 and group 4 examinees, the false positives and the false negatives. Raising the standards would prevent people without the ability from entering the profession, but at the cost of also keeping out group 4 examinees, with the skill but with low test scores. Lowering the passing score will protect group 4 examinees, but at the cost of admitting more group 3 examinees, those with adequate scores but without the ability.

Standards Established for NTE Tests of Basic Skills

As shown in Table 1, 12 out of 14 states using the NTE adjusted the passing scores to be lower than the study scores. Only North Carolina and Rhode Island chose to use the score the state panel expected to be obtained by a minimally qualified individual. The passing scores range from a low of 630 to a high of 657, and average 8 points less than the study scores.

These are relatively large adjustments. The standard errors of measurement range for the Communication Skills, General Knowledge, and Professional Knowledge tests are 3.5, 3.5 and 3.8, respectively. The average adjustment, then, is two standard errors of measurement downward.

In order to examine the impact of these adjustments, the distributions of NTE Professional Knowledge tests scores was examined. The percentile rankings corresponding to each scaled NTE score (NTE, 1984) were converted to interval frequencies. This distribution along with the passing score and the study score are shown in Figure 1. Graphs for the other tests show similar adjustments and score distributions.

Table 1. Study Scores, Actual Passing Scores, and Pass Rates for States Using the NTE in 1987

State	Commun. Skills			General Knowledge			Prof Knowledge			Pass Rate
	study	actual	diff	study	actual	diff	study	actual	diff	
Idaho	656	652	-4	650	646	-4	656	648	-8	n/a
Indiana	659	653	-6	655	645	-10	640	646	6	88
Kansas							645	642	-3	94*
Kentucky	663	643	-20	658	637	-21	661	641	-20	93
Louisiana	652	645	-7	651	644	-7	652	645	-7	87
Mississippi	652	644	-8	647	639	-8	650	642	-8	88
Montana	652	648	-4	648	644	-4	652	648	-4	92
New Jersey				656	646	-10				83
New Mexico	656	644	-12	657	645	-12	642	630	-12	88*
New York	656	650	-6	656	649	-7	652	646	-6	79*
North Carolina							644	644	0	80*
Rhode Island	657	657	0	649	649	0	648	648	0	n/a
Tennessee	662	644	-18	658	640	-18	655	635	-20	n/a
Virginia	651	649	-2	641	639	-2	641	639	-2	n/a
minimum	651	643	0	641	637	0	640	630	6	79
maximum	663	657	-20	658	649	-21	661	648	-20	94
mean	656	648	-8	652	643	-9	649	643	-6	87

* denotes passing rate on the most difficult test not the overall pass rate.
n/a denotes information not available.

Source: A.G. Zetler "Montana Validation of the NTE Core Battery: Study Report," Contractor's report, January 1986; L.M. Rudner (ed) "What's Happening in Teacher Testing" Washington, DC: Government Printing Office, Office of Educational Research and Improvement, 1987; T.E. Eissenberg L.M. Rudner "Teacher Testing: The 1988 Report" (1988).

The intent of the adjustments is to reduce the probability of erroneously rejecting a minimally qualified applicant. To further illuminate the effect of these adjustments, two groups of examines were identified.

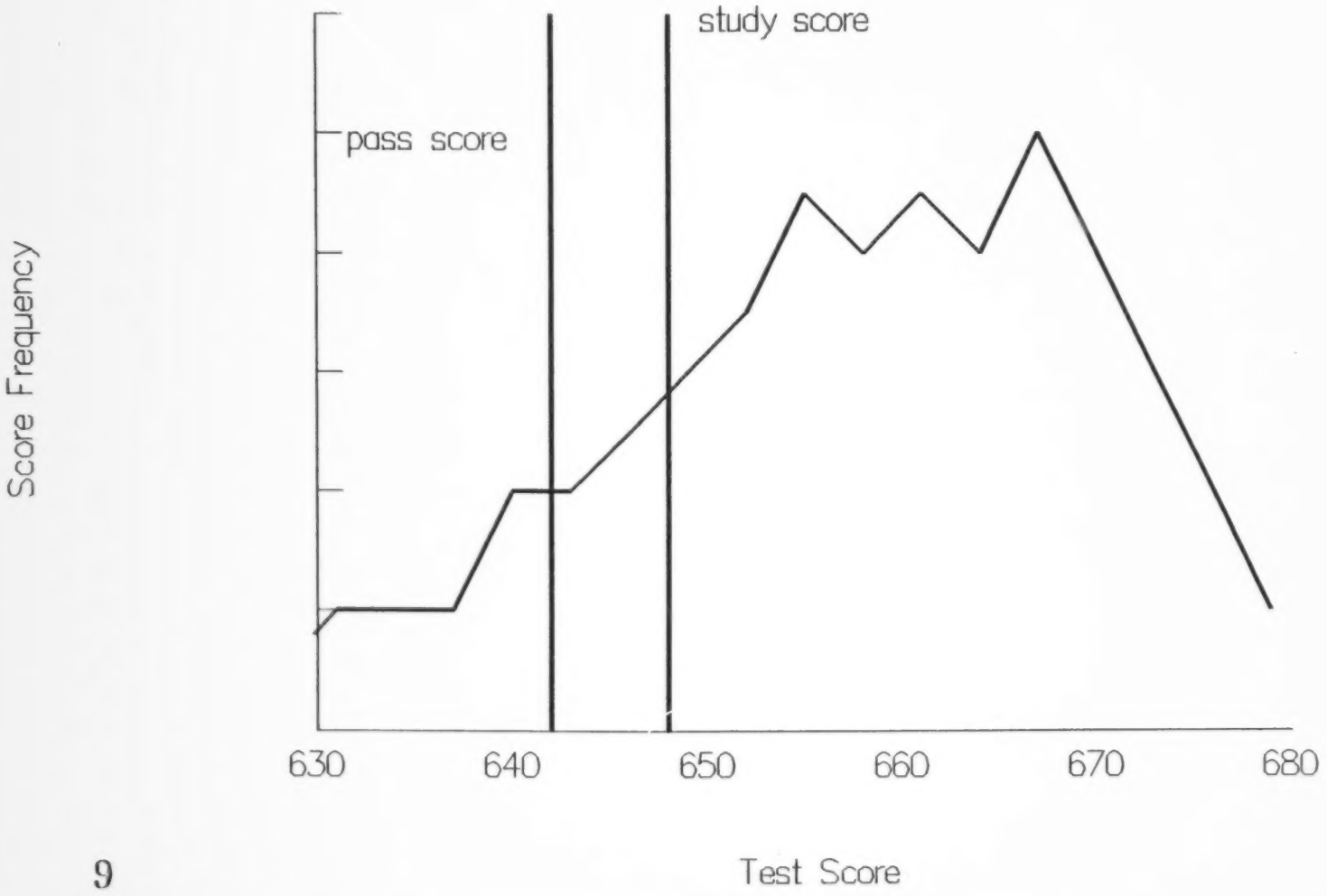
1) marginally unqualified individuals whose observed scores were one standard error of measurement (4 points) below the study score, and
2) marginally qualified individuals whose observed scores were one standard error of measurement above the study score.

With an 8-point adjustment, there is a less than 1 in 1,000 chance of rejecting these marginally qualified applicants. The adjustments, however, also greatly increase the probability of accepting an unqualified applicant. With an 8-point downward adjustment, there is now an 8 in 10 chance of accepting an marginally unqualified applicant. The 7- to 9-point downward adjustments from the average study scores raise the pass rates from approximately 66 percent, 61 percent, and 74 percent on the Communication Skills, General Knowledge, and Professional Knowledge tests to approximately 82 percent, 81 percent, and 88 percent, respectively.

With these downwardly adjusted passing scores, the actual number of items a teacher candidate needs to answer correctly in order to pass the examinations is relatively low. The 1982 Professional Knowledge test, for example, is comprised of 104 questions. The average passing score set by states using the test was 642. To obtain this score, one only needs to answer 47 of the 104 items correctly. Passing scores range from 35 items in the state with the lowest passing score to 53 items in the state with the highest passing score. This is not much higher than the chance level. The test is scored on the basis of the number of correct answers. No points are subtracted for incorrect answers or omitted questions. With 5 choices for each item, a candidate should be able to answer 21 questions correctly by randomly marking the answer sheet.

Had study scores, rather than adjusted scores, been used as the passing scores, passing rates would have been considerably lower. Based on the national distribution of scores, the passing rates would have been 12 to 20

Distribution, Study Score & Pass Score for the NTE Professional Skills Test



percent lower on each of the three tests. Multiplying the number of newly certified individuals in these 14 NTE states by 16 percent indicates that possibly 11,000 candidates in these 14 states have scores in the safety range between the study score and the actual cutoff score used by the state. Since the teacher turn-over rate is approximately 20% (Feistritzer, 1985), some 3-4% of the teaching work force may be in this marginal range.

Discussion

The resurgence of teacher testing during in the late 1970's and 1980's, began in an era of increased criticism of the schools, declining student test scores, the student competency testing movement, and most importantly, when teacher training programs were producing almost twice as many prospective teachers as there were openings. But these programs take time to implement. The state legislature usually debates the idea and gathers information well before taking draft legislation through to laws and regulations. State departments of education spend years planning the program. Instruments are researched or developed. Validation studies are conducted. Vast amounts of advice and public input is obtained. As a result, states are just beginning to implement mandates that were conceived at the start of the decade.

We have now entered into an era where schools are getting better. We have also entered into an era where we are faced with a shortage of teacher education students and teacher education graduates. Testing programs originally designed in the name of increased standards and tougher access to the profession are now out of sync with the times. Many schools, especially those in urban areas, need warm adult bodies to fill their classrooms. Schools must weigh the need for warm bodies against what the tests are capable of doing. Naturally, the need for warm bodies are winning. Passing criteria for many state teacher licensure testing programs are quite low. Approximately 83% of those taking teacher licensing tests pass the first time. With virtually everyone passing, are these current programs, worth the time, expense, and aggravation they incur?

At best, these programs are able to weed out only grossly incompetent teacher candidates. Raising the passing criteria and hence lowering the passing

rate, however, is not the solution. This would only exacerbate the teacher shortage. The original issues of quality and public confidence would still not be addressed.

Many teacher tests cover minimal academic skills that most people acquire by eighth grade. People who cannot pass a simple test of basic knowledge should, a priori, not be placed in a position where they are responsible for the education of children. Just as this is self-evident, it is also naive to expect a basic skills test to serve as a meaningful standard or to enhance the teaching profession. If anything, such a test is an affront to the professionalism it supposedly establishes. In Texas, for example, a test of eighth grade ability was administered as that state's recertification test. Teachers were outraged; they considered the testing program an embarrassment and an insult (Shepard and Kreitzer, 1987).

Conclusions

The stated and apparent goals of teacher certification testing programs are impressive. Advocates claim they improve the caliber of school teachers, promote excellence, attract better teachers, and assure the public that students receive a quality education. The practical side of testing, which must take into account state finances, time, the current state of the art in psychometrics, supply and demand, and political realities, however, appears to preclude high standards.

The average passing score on the Professional Knowledge examination of the NTE is 47 out of 104 items -- a value that is not much higher than chance. The average passing scores on the other NTE examinations are equally low. The states have typically adopted standards that are much lower, approximately 2 standard errors of measurement lower, than the cut scores recommended by advisory panels.

The disjuncture between the rhetoric and current standards can have serious implications. Acceptance of the rhetoric could easily lead to a sanctioning of these low standards and preclude the use of more relevant and more rigorous testing programs.

REFERENCES

American Association for Colleges for Teacher Education (1986), Teacher Education for the United States, Washington, DC: author.

NTE (1984) "Interpreting Scores on the NTE Core Battery Tests, 1984-85", Information flier published by the NTE Program of Educational testing Service, Princeton, NJ.

Eissenberg, T.E. and L.M. Rudner (1988) "Teacher Testing: The 1988 Report", ERIC/TM Digest 88-1, Washington, DC: American Institutes for Research.

Feistritzer, C.E (1985) The Condition of teaching, A state by state analysis, 1985. Washington, DC: The Carnegie Foundation for the Advancement of Teaching.

Rudner, L.M. (ed, 1987) What's Happening in Teacher Testing, Washington, DC: Government Printing Office, Office of Educational Research and Improvement.

Rudner, L.M. (Spring, 1988) "Teacher Testing - An Update", Educational Measurement: Issues and Practice, 7(2), in press.

Shepard, L.A. and A.E. Kreitzer (August-September, 1987) "The Texas Teacher Test", Educational Researcher, 16 (6), 22-31.

Zetler, A.G. (1986) "Montana Validation of the NTE Core Battery: Study Report", Contractor's report.